An invariant Arginine in common with MHC class II allows extension at the C-terminal end of peptides bound to chicken MHC class I

An invariant arginine in common with major histocompatibility complex (MHC) class II molecules allows extension at the C-terminal end of peptides bound to chicken MHC class I molecules

Running title: Chicken MHC class I with an open groove.

Jin Xiao [*,†,1], Wangzhen Xiang [*,†,1], Yongli Zhang[‡], Weiyu Peng[§,†], Min Zhao [¶], Ling Niu [¶], Yan Chai [¶], Jianxun Qi [¶], Fei Wang[*], Peng Qi[*], Chungang Pan[*],Lingxia Han[‖], Ming Wang [*,†], Jim Kaufman [#,**], George F. Gao[‡, § ,¶ , ††] , William J. Liu[‡, §]

* Key Laboratory of Veterinary Bioproduction and Chemical Medicine of the Ministry of Agriculture, Engineering and Technology Research Center for Beijing Veterinary Peptide Vaccine Design and Preparation, Zhongmu Institutes of China Animal Husbandry Industry Co. Ltd, Beijing 100095, China

† College of Veterinary Medicine, China Agricultural University, Beijing 100193, China

‡ College of Laboratory Medicine and Life Sciences, Institute of Medical Virology, Wenzhou Medical University, Wenzhou 325035, China;

§ Key Laboratory of Medical Virology, Ministry of Health, National Institute for Viral Disease Control and Prevention, Chinese Center for Disease Control and Prevention, Beijing 100052, China;

¶ CAS Key Laboratory of Pathogenic Microbiology and Immunology, Institute of Microbiology, Chinese Academy of Sciences, Beijing 100101, China;

‖State Key Laboratory of Veterinary Biotechnology, Harbin Veterinary Research Institute, Chinese Academy of Agricultural Sciences, Harbin 150001,China;

#Department of Pathology, University of Cambridge, Cambridge CB2 1QP, United Kingdom;

**Department of Veterinary Medicine, University of Cambridge, Cambridge CB2 1QP, United Kingdom;

†† China Research Network of Immunity and Health (RNIH), Beijing Institutes of Life Science Chinese Academy of Sciences, Beijing 100101, China.

Footnotes

[1] JX and WX contributed equally to this work.

[2] Corresponding: **Dr. William J. Liu**, National Institute for Viral Disease Control and Prevention, Chinese Center for Disease Control and Prevention, Beijing 100052, China. Phone: +86-10-63516568, Fax: +86-10-63510565, Email: liujun@ivdc.chinacdc.cn; or **Dr. George F. Gao**, National Institute for Viral Disease Control and Prevention, Chinese Center for Disease Control and Prevention, Beijing 100052, China. Phone: +86-10-64807688, Fax: +86-10-64807522. Email: gaof@im.ac.cn; or Dr. Jim Kaufman, Department of Pathology, University of Cambridge, Tennis Court Road, Cambridge CB2 1QP, UK. Email: jfk31@cam.ac.uk.

[3] This work was supported by grants from the National Natural Science Foundation of

China (Grants 81401312 and 81373141). G.F.G. is a leading principal investigator of the National Natural Science Foundation of China Innovative Research Group (Grant 81621091).The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Summary

MHC molecules are found in all jawed vertebrates, and are known to present peptides to T lymphocytes. In mammals, peptides can hang out either end of the peptide-binding groove of classical class II molecules, whereas the N- and C-termini of peptides are typically tightly bound to specific pockets in classical class I molecules. The chicken MHC, like many non-mammalian vertebrates, has a single dominantly-expressed classical class I molecule, encoded by the BF2 locus. We determined the structures of BF2*1201 bound to two peptides, and found that the C-terminus of one peptide hangs outside of the groove, with a conformation much like the peptides bound to class II molecules. We found that BF2*1201 binds many peptides that hang out of the groove at the C-terminus, and the sequences and structures of this MHC I allele was determined to investigate the basis for this phenomenon. The classical class I molecules of mammals have a nearly invariant Tyr (Tyr84 in humans) that coordinates the peptide C-terminus, but all classical class I molecules outside of mammals have an Arg in that position, in common with mammalian class II molecules. We find that this invariant Arg residue switches conformation to allow peptides to hang out of the groove of BF2*1201, suggesting that this phenomenon is common in chickens and other non-mammalian vertebrates, perhaps allowing the single dominantly-expressed class I molecule to bind a larger repertoire of peptides.

Introduction

The major histocompatibility complex (MHC) is a genetic region found in all vertebrates, encoding class I and class II molecules with crucial roles in immune responses (1, 2). The classical molecules are highly polymorphic, specifically binding certain peptides depending on the allele, and then presenting them for recognition by T lymphocytes. Although there are exceptions and other important mechanisms, broadly speaking the classical class I molecules present peptides derived from the cytoplasm and nucleus (where viruses and a few bacteria replicate) to CD8-bearing cytotoxic T lymphocytes (CTLs), while classical class II molecules present peptides derived from intracellular vesicles and the extracellular space (where bacteria, parasites and fungus are found) to CD4-bearing regulatory and effector T cells. In addition to the adaptive immune system, the classical class I molecules are recognized by natural killer (NK) cells as part of the innate immune system (2, 3).

The groove of class II molecules is open on either end, so that the peptides can vary in length, with the amino acids outside the nonamer core hanging out the ends (3, 4). In contrast, the groove of classical class I molecules are blocked at each end, so that peptides longer than octamers bulge up in the middle. The peptides bound to class I molecules are generally short (8-10mers, although much longer peptides have been described) with free N- and C-termini (3, 5, 6). The free N- and C-termini are co-ordinated by eight invariant residues, including Tyr84 that binds the C-terminal

carboxyl group (7-10). Until recently, no extensions at the N- or C-termini had been reported, but a C-terminal extension was reported to reduce binding affinity by ten-fold (11).

In addition to the conserved binding at the N- and C-termini, specificity of binding to class I molecules is largely determined by anchor residues that generally have one or a few amino acids with similar chemical characteristics, resulting in a characteristic motif for each class I molecule (5). Anchor residues have been found at a number of positions (along with other positions that have preferred residues, often referred to as secondary anchors), but most typically mammalian class I molecules have anchor residues at peptide position 2 (P2) and at the C-terminal position (PΩ), which fit into so-called pockets B and F, respectively (5). One well-studied exception is the mouse class I molecule H-2K^{b}, which has anchor residues at PΩ, and P5 for octomers (P6 for nonamers), with the particular amino acid found at P2 reported to co-vary with the amino acid at P5(P6) (9, 12, 13).

Compared to mammals, the chicken MHC is simple and compact, with only the class I molecule encoded by the BF2 locus expressed at a high level (14-16). The reason for this dominantly-expressed class I molecule is the fact that the chicken MHC, in stark contrast to the MHC of typical mammals, suffers very little recombination and encodes polymorphic peptide-loading complex (PLC) molecules, including transporters

116 associated with antigen processing (TAPs) and tapasin, which for each haplotype
117 co-evolve with the particular BF2 allele present (17-20).
118
119 In comparison to the MHC of typical mammals, the chicken MHC (the BF-BL region
120 or the B locus in which it is located) determines striking genetic associations with
121 resistance or susceptibility to infectious diseases, many of which are of economic
122 importance for the global poultry industry (21, 22). The multigene family of class I
123 molecules expressed by typical mammals means that there are multiple chances to
124 present a protective peptide, so most MHC haplotypes confer more-or-less resistance to
125 most pathogens, which reads out as weak genetic associations. In contrast, the single
126 dominantly-expressed class I molecule of a chicken MHC haplotype will either find a
127 protective peptide or not, so a particular MHC haplotype can determine life or death for
128 a chicken, which reads out as strong genetic associations (23, 24). Thus, the peptide
129 motif for the dominantly-expressed class I molecule is particularly important. These
130 motifs range from those much more promiscuous (BF2*0201, 1401 and 2101) to those
131 more fastidious (BF2*0401, 1201, 1501 and 1901) than those reported for humans and
132 mice (15, 25-27). The peptides determined for the fastidious class I molecules have
133 been largely octamers with three anchor residues, with a structure reported for
134 BF2*0401(26).
135
136 The first chicken class I gene identified encodes the fastidious BF2*1201 molecule
137 (now officially designated BF2*012:01), for which the peptide motif was reported

138 generally to be octamers with Val or Ile at position 5, and Val at position 8 (VI5-V8)
139 (15, 28, 29). This original motif as well as less stringent versions were used for
140 prediction and testing of T cell epitopes for a variety of chicken pathogens and
141 vaccines, including Rous sarcoma virus (RSV), infectious bursal disease virus (IBDV)
142 and avian influenza virus (AIV) (15, 30, 31). In the first and most complete analysis, 13
143 peptides were predicted for the *v-src* gene of RSV Prague strain C, for which chickens
144 of the B12 haplotype were resistant. Of these putative T cell epitopes, only three were
145 found to increase the amount of class I molecules stabilized at the cell surface, and
146 immunization with the strongest binding peptide was found to reduce incipient tumors
147 (15, 32). Similarly, peptides were predicted for an IBDV molecular vaccine that
148 correlated with protection (30), and for AIV infection that elicted interferon-γ
149 production by CD8 T cells (31). Thus, the peptide motif for BF2*1201 is of interest for
150 explaining disease resistance and vaccine responses.
151
152 Two additional mysteries arise from the original study of BF2*1201(15). First, no
153 obvious similarities were found for the amino acids at P2, and no evidence for the kind
154 of co-variation reported for H-2K^{b} in mouse was noted. Thus, the role of P2 residues
155 binding into pocket B remains unknown. Second, there was one peptide eluted from
156 B12 cells that was a nonamer with a Val at P8 and a Thr at P9. One possibility was that
157 the methyl group of the Thr binds into pocket F like a methyl group of Val in other
158 peptides, but another possibility was that Val8 fits into pocket F and Thr9 is outside the
159 groove. Several recent studies indicate that the C-terminal-extended peptides can bind

to some mammalian MHC I alleles upon significant conformational shifts, including the Tyr84 (49, 50, 61). Interestingly, the sequences of chicken class I molecules have seven of the eight invariant residues that co-ordinate the N- and C-termini of the peptides, but the equivalent of the Tyr84 is always an arginine in chickens, and indeed in virtually all non-mammalian vertebrates (33, 34). Moreover, for class II molecules that allow the peptides to hang out the end of the groove, the equivalent position is also an arginine (33, 34). Thus, the question arises whether this arginine might allow peptides to hang out of the C-terminal end of the peptide-binding groove.

The salient features of the chicken MHC and class I molecules are found in many if not most non-mammalian vertebrates (35, 36). Thus, the chicken provides a model for these other vertebrates, some of which are of economic or conservation interest. In addition, there are there several mysteries about BF2*1201, so we decided to examine the structure and peptides further.

Materials and Methods

Peptide synthesis and preparation of expression constructs. Potential chicken MHC I BF2*1201-binding peptides and their substitutions (Table 1, TableS1) were synthesized and purified by reverse-phase high-performance liquid chromatography (HPLC) (SciLight Biotechnology, Beijing). The peptide purity was determined to be >95% by analytical HPLC and mass spectrometry. The peptides were stored at -80°C as freeze-dried powders and were dissolved in dimethyl sulfoxide (DMSO) before use. The cDNAs for BF2*1201 (GenBank: Z54329.1) were synthesized (Genewiz Inc, Beijing). The amplified products expressing the extracellular domain (residues 1–270) of BF2*1201 were cloned into a pET21a vector (Novagen) and transformed into Escherichia coli strain BL21(DE3). The expression plasmid for chicken β2m (expressing residues 1-98) was constructed previously in our laboratory (37).

Refolding and purification of BF2*1201. Dilution-renaturation and purification of MHC I assembled with peptides were performed as described previously (37). Firstly, 1 mL dissolved chβ2m inclusion bodies was dropped slowly to 500 mL refolding buffer (100 mM Tris-HCl pH 8.0, 2 mM EDTA, 400 mM L-Arg, 0.5 mM oxidized glutathione, 5 mM reduced glutathione) and incubated at 4℃for 0.5 h. Subsequently, 5 mg peptide dissolved in dimethyl sulfoxide (DMSO) was added to the solution. Half an hour later, 3 mL BF2*1201 heavy chain inclusion bodies were added to the solution drop by drop. After incubation for 8 h, the soluble portion was concentrated and

197 purified by chromatography on a Superdex 200 16/60 HiLoad (GE Healthcare)
198 size-exclusion column. For the crystallization, the protein was further purified by
199 Resource Q (GE Healthcare) anion-exchange chromatography.
200 Crystallization, data collection, and processing. Crystallization of was performed using
201 the sitting drop vapor diffusion technique. Plates were incubated at 291 K and assessed
202 for crystal growth after 1-2 weeks. BF2*1201/S1 crystals were observed in 0.1M
203 BICINE pH 8.5,20% w/v PEG 10,000 at a concentration of 16mg/mL. Single crystals
204 of BF2*1201/A6 were grown in10%v/v polyethylene glycol 200, 0.1M BIS-Tris
205 propane pH 9.0, 18%w/V, Polyethylene glycol 8,000 at a protein concentration of 16
206 mg/mL. Diffraction data for both crystals were collected at 100 K at the SSRF
207 BEAMLINE BL17U, Shanghai, China at a wavelength of 0.97914 Å and 0.97539 Å,
208 respectively. In both cases, the crystals were first soaked in reservoir solution
209 containing 15% glycerol as a cryoprotectant for several seconds and then flash-cooled
210 in a stream of gaseous nitrogen at 100 K. The collected intensities were subsequently
211 processed and scaled using the DENZO program and the HKL2000 software package
212 (HKL Research)(38).
213
214 Structure determination and analysis. The structure of BF2*1201/A6 was determined
215 by molecular replacement using BF2*0401/IE8 (Protein Data Bank [PDB] code 4E0R)
216 as a search model in the Crystallography & NMR System (CNS) program (39). Clear
217 solutions in both the rotation and translation functions correspond to two molecules in
218 one asymmetric unit. The residues that were different in BF2*1201 and the search

219 model BF2*0401/IE8 were manually rebuilt in the COOT program under the guidance
220 of Fo-Fc and 2Fo-Fc electron density maps (40). Subsequently, we refined the initial
221 rigid body and performed a series of restrained translation, libration, and screw-motion
222 (TLS) refinements with the REFMAC5 program (41). Additional rounds of refinement
223 were performed using the phenix refine program implemented in the PHENIX package
224 with isotropic ADP refinement and bulk solvent modeling (42). We assessed the
225 stereochemical quality of the final model with the PROCHECK program (43). The
226 structure of BF2*1201/S1 was determined by using the final model of BF2*1201/A6
227 through molecular replacement, with four molecules present in one asymmetric unit.
228 The same refinement steps were performed to generate the final model of
229 BF2*1201/SRC. Crystallographic statistics for the final models of the two complexes
230 are given in Table S1. The molecular models and the electron density-related figures
231 were generated using PyMOL (http://www.pymol.org/). The sequence alignment was
232 generated with Clustal X (44) and ESPript (45).
233
234 Thermostability measurements using circular dichroism (CD). To evaluate the
235 thermostability of MHC I/ peptide complexes, we used the CD spectroscopy method as
236 previously described (46). All complexes were prepared as described above and diluted
237 to 200 μg/mL in 20 mM Tris-HCl (pH 8.0) and 50 mM NaCl. CD spectra at 218 nm
238 were measured on a Chirascan spectrometer (Applied Photophysics) using a
239 thermostatically controlled cuvette at temperature intervals of 0.2°C and a rate of
240 1°C/min between 20-90°C. The proportion of denatured protein was calculated from

241 the mean residue ellipticity (u) using the standard method: fraction unfolded (%) = (θ
242 −θa)/(θa−θb), where θa and θb are the mean residue ellipticity values in the fully
243 folded and fully unfolded states, respectively. The denaturation curves were generated
244 by nonlinear fitting with OriginPro 8.0 (OriginLab). The midpoint transition
245 temperature (Tm) was calculated by fitting data to the denaturation curves and using
246 inflexion-determining derivatives.
247
248 Accession numbers. Atomic coordinates and structure factors have been deposited in
249 the Protein Data Bank (http://www.rcsb.org) under accession codes 5YMW for
250 BF2*1201/S1 and 5YMV for BF2*1201/A6.
251

Results

Immunogenic peptides from pathogens based on the motif (V/I)5-V8 will allow refolding of BF2*1201

Peptides eluted from class I molecules on B12 cells suggested a motif of (V/I)5-V8, which was consistent with a structural model of the BF2*1201 molecule (15). This motif (relaxed to allow nonamers as well) predicted many peptides from the sequence of RSV Prague strain C, and three peptides from the *v-src* gene gave a detectable signal in a cell-binding assay, which here we call B12-P1 (GENLVCKV, with putative anchor residues underlined), B12-P2 (LAGGVTTFV) and B12-S1 (LPACVLEV). Immunization with B12-S1 protected B12 chickens from tumors after RSV infection (32). The same motif (relaxed to allow nonamers as well as related hydrophobic amino acids) was used to predict peptides from AIV sequences (A/chicken/Italy/1067/99), from which we chose to examine five peptides, which here we call B12-A1 (NATEIRASV), B12-A6 (AVKGVGTMV) and B12-A11 (EDLRVSSFI) from the nucleoprotein, and B12-F1 (VETYVLSI) and B12-F7 (KTRPILSPL) from matrix protein 1. These peptides were shown to stimulate T cells from infected chickens to produce interferon-γ (31). However, no careful molecular assay has been used to examine the binding of these peptides to the BF2*1201 molecule (Table I).

Using a refolding assay (renaturation followed by size exclusion chromatography) as in previous work (25, 26), we found that six peptides supported refolding to give a stable

monomer (peptide, heavy chain and β_2m) of the BF2*1201 molecule (Fig. 1). Five peptides with the canonical motif ((V/I5)-V8 for octamers, (V/I)5-V9 for nonamers) gave good yields, one octamer with a non-canonical residue in the PΩ position (V5-I8) gave a lesser yield, and two nonamers with non-canonical Ile and Leu residues at PΩ gave no properly folded protein. The results are clear for these few peptides, and can be interpreted to suggest that octamers are the most stable length and therefore may tolerate non-canonical residues in some anchor positions. This interpretation is consistent with the self peptides, which were octamers with (V/I)5-V8, except one with V5-L8 and one nonamer.

For the six peptides that supported refolding, several amino acids were found at P2: two with Ala, one with Val, one with Pro, and two with Glu (Table I). This result is consistent with the self peptides eluted from B12 cells, for which there were two with Ala, two with Gln and one with Thr (15). Both for self peptides and the six confirmed predicted peptides, there was no evidence for one or a few amino acids with similar chemical properties at P2, nor any evidence for co-variation of position P2 with position P5 as is found in H-2K^b.

Two structures of BF2*1201 with pathogen peptides show three positions that anchor the peptide

297 The fact that a variety of amino acids, ranging from small hydrophobic and small polar
298 to large polar and negatively-charged residues, are found at P2 of BF2*1201-binding
299 peptides leaves open the question whether P2 is an anchor residue in the classic sense.
300 In order to answer this question, and to further understand structure-function
301 relationships, we determined the crystal structures of BF2*1201 in complex with the
302 RSV-derived peptide S1 and the AIV-derived peptide A6 (Table II).
303
304 Overall, the structure of BF2*1201 has the common characteristics of other classical
305 class I molecules: the extracellular region of the heavy chain folds into three different
306 domains (Fig. S1A), the α1 and α2 domains form a typical peptide binding groove
307 which contains two α1-helices and eight β-sheets, and the α3 domain and β_2m display
308 typical immunoglobulin domains and underpin the peptide binding domain. The
309 superposition of BF2*1201 onto previously determined BF2 alleles from the B2, B4,
310 B14 and B21 haplotypes showed root mean square deviations (RMSDs) of 0.55 to 0.71
311 Å (Fig. S1B). The most distinct differences between BF2*1201 and the other alleles are
312 located in the loop between β1 and β2 strands of the α2 domain and the β1-β2 loop in
313 the α3 domain (Fig. S1B).
314
315 As expected for octamer peptides bound to class I molecules, the peptide S1 and the
316 first eight amino acids of peptide A6 are flat and low in the peptide-binding groove
317 (Fig. 2A, B), with the Val5 in both peptides located as deep in the groove as the amino
318 acids at P2 and P8 (Fig. 2C, D). Compared to the crystal structures of octamer peptides

bound to various class I molecules from chickens, mice and humans, the Cα atom of the P5 residues are located 0.84 Å lower than the lowest previously reported (Fig. 3).

Like the fastidious chicken allele BF2*0401 (now officially designated as BF2*004:01) (26), BF2*1201 has a narrow groove but in contrast, the BF2*1201 groove is overall hydrophobic as revealed by the vacuum electrostatic surface potential (Fig. 2E, F). As expected, the amino acids at P5 (Val in both peptides) and P8 (Val in S1 and Met in A6) are anchor residues located in the C and F pockets respectively, but the amino acids at P2 (Pro in S1 and Val in peptide A6) also act as anchor residues, being located deep in the B pocket, nestling under the α1 helix. For these peptides, the anchor residues have relatively small hydrophobic side chains, but larger polar and charged amino acids like Glu and Gln have also been found at P2 (15). The way in which these larger side chains are accommodated remains to be determined.

In order to examine the role of these anchor residues more closely, Gly swaps for the three positions were tested by refolding assays, and the thermal stability of resulting monomers was monitored by circular dichroism (CD) spectroscopy (Fig. 4, Table S1). Substitution of Gly at P8 led to no refolding for either peptide, and Gly at P5 in the AIV peptide A6 also led to no refolding, whereas Gly at P5 in the RSV peptide S1 supported refolding although the monomer was significantly less stable than the original peptide. In contrast, substitution of Gly at P2 in either peptide led to only minor reductions in the yield of refolded monomers, which had similar stabilities compared to the original

peptides. So, by these analyses, the P2 anchor of BF2*1201 binding peptides seems more promiscuous compared to the P5 and P8 anchors.

The stability of peptide binding to BF2*1201 is not affected by C-terminal residues hanging out of the groove

By far the most common C-terminal anchoring mode, found in nearly every structure of mammalian class I molecules, is for the side chain of the amino acid at PΩ to insert into pocket F, and the C-terminal carboxyl group to bind with a hydrogen bond to Thr141 and Tyr84. Examples include the SARS-CoV peptide Mn2 bound to HLA-A*0201 and the MERS-CoV peptide 37-1 bound to H-2K^{d} (Fig. 5C-F) (47, 48). A similar mode of binding is found for the RSV octamer peptide S1 (and all other structures of chicken class I molecules known), although Tyr84 in mammals is replaced by Arg83 in chickens (Fig. 5A, B). However, most striking observation from the structures of BF2*1201 is the fact that the anchor residue for the AIV nonamer peptide A6 is Met8, with the Val9 (which was expected to be the anchor residue) hanging out over the F pocket. The peptide bond between Met8 and Val9 makes a hydrogen bond with Thr141, but the carboxyl group of Val9 makes the hydrogen bond with Arg83, pushing the side chain of Arg83 away to create a more open conformation (Fig. 5G, H).

Recently, two structures of mammalian class I molecules were reported that also have C-terminal amino acids hanging out of the groove (49, 50): *Toxoplasma gondii* peptide FK12 (FVLELEPEWTVK) bound to HLA-A*0201 (Fig. 5I, J; Fig. S2B) and insulin peptide G9GF (LYLVCGERGF) bound to H-2K^{d} (Fig. 5K, L; Fig. S2C). In both cases, a single amino acid at the peptide C-terminus hangs out of the groove and Tyr84 is shifted away, but in neither case does the amino acid at PΩ interact with Tyr84. We compared these structures to HLA-A*0201 and H-2K^{d} bound to peptides with the common C-terminal anchoring mode, and found that the α-helices were significantly shifted (Fig. 5N, O), with RMSD of 0.402 Å for HLA-A2 and 0.588 Å for H-2K^{d}. In contrast, the α-helices of BF2*1201 with peptide S1 and A6 were not shifted (Fig. 5M), with the RMSD of 0.225 Å not significantly different from RMSD of 0.201 Å between the two molecules in the asymmetric unit of the BF2*1201/S1 structure.

We tested whether a C-terminal residue hanging out of the groove would affect stability of different MHC molecules. The AIV nonamer peptide A6 (AVKGVGTMV) and the octamer A6-N8 (AVKGVGTM, the same peptide with the last amino acid deleted) refolded equally well with BF2*1201, and the resulting monomers were equally stable (Fig. 6A, B). In contrast, HLA-A*0201 refolded poorly and gave less stable monomers with the peptide P9-C10 (AIMEKNIVLK) extended one amino acid compared to the parent peptide P9 (AIMEKNIVL, HLA-A2 restricted peptide from 2009 pandemic H1N1 influenza virus(51))(Fig. 6C, D). Similarly, HLA-A*1101 didn't refold at all with the peptide P23-C11 (RFSSFIRGKKV) extended one amino acid compared to the

parent peptide P23 (RFSSFIRGKK, HLA-A11 restricted peptide from 2009 pandemic H1N1 influenza virus(51)) (Fig. 6E, F). These results for mammals fit well with the original observation by Wiley and colleagues (11) that a decamer calreticulin peptide with a single amino acid C-terminal extension bound to HLA-A*0201 has a marked effect on stability of binding.

Finally, we tested by refolding the AIV peptide S1 (31) as well as five self peptides eluted from B12 cells (15), here called P4 – P8, extended by up to three C-terminal amino acids. All of the extended peptides refolded as well as the parent peptide, indicating a general ability of BF2*1201 to accommodate peptides with amino acids that hang out of the groove at the C-terminal end (Table S1). To further investigate whether the Arg at the C-terminus of the peptide binding groove is responsible for the intrinsic binding of the C-terminal-extended peptides by BF2*1201, we generated a R83Y mutant of BF2*1201, which has a Tyr at the position 83. Interestingly, R83Y mutant of BF2*1201 lost the capacity to bind the AIV nonamer peptide A6 (AVKGVGTMV), indicating a key role for Arg83 in the binding of C-terminally-extended peptides (Fig. 7).

The classical class I molecules of non-mammalian vertebrates generally have Arg83, which is in the same relative position as an Arg in mammalian class II molecules

406 The first report of the sequence and molecular model of BF2*1901 (now officially
407 designated BF2*015:02) pointed out that one of the eight invariant residues identified
408 by Wiley's group as coordinating the N- and C-termini of the bound peptide, Tyr84 in
409 humans, is replaced by Arg83 in chickens, and that this position also is an Arg in human
410 class II molecules (33). As more sequences became available, it became clear that the
411 classical class I molecules of all placental mammals had Tyr while all known
412 non-mammalian vertebrates had Arg (34).
413
414 We revisited this analysis by downloading class I sequences from many vertebrates.
415 We find that, with very few exceptions, the classical class I molecules from almost all
416 mammals, including marsupials and monotremes (52, 53), have Tyr in the position
417 equivalent to the human Tyr84, while representative birds, reptiles, amphibians, bony
418 fish and cartilaginous fish all have Arg (Fig. 8; Table III). Based on our finding with
419 BF2*1201, the presence of this Arg suggests that class I molecules throughout the
420 non-mammalian vertebrates have the capacity to accommodate peptides which extend
421 beyond pocket F.
422
423 The equivalent position in mammalian class II α chain sequences is also an Arg (Fig.
424 9A, Table III), and structures reveal that this Arg forms a hydrogen bond with the
425 carbonyl of the peptide bond between the last anchor residue and the first amino acid
426 that hangs out of the groove (Fig. 9E-I). The conformation of peptides hanging out of
427 the groove is similar in class II molecules from humans and mice, and to the chicken

class I molecule BF2*1201 (Fig. 9B, C). However, while the conformation of the Arg in class II molecules is similar in all structures, the conformation of this Arg in the chicken class I molecule BF2*1201 is completely different for the overhanging peptide. For the peptide that is not hanging out of the groove, the conformation of this Arg in BF2*1201 is very similar to the other structures for chicken class I molecules, and to canonical conformation of the equivalent Tyr in mammalian class I molecules (Fig. 9C, D). The presence of this Arg correlates perfectly with the ability of class I and II molecules to allow peptides to hang out of the groove.

The classical class I molecules in chickens have Asp14 which restrains a loop interaction with β_2m

Examining the sequences of classical class I molecules from a variety of vertebrates identified another position which is shared among mammals, but is different among non-mammalian vertebrates. In the position equivalent to Arg14 found in virtually all mammalian sequences including marsupials and monotremes (52, 53), the sequences from birds, reptiles and amphibians have an Asp (Fig. S3A). In structures of chicken class I molecules, Asp14 forms a salt bridge with Lys33 of β_2m, resulting in a consistent conformation of the loop between the first and second β-strands of the α1 domain, and a close interaction with β_2m (Fig. S3B, C). In *Xenopus*, the equivalent residue in β_2m is Arg34 (54), which presumably can make the same interaction. By

contrast, the structures of mammalian class I molecules show multiple conformations of this loop, all of which are farther away from β_2m which has various residues at position 33 including Asp, Gln and His (55). Thus, other important positions besides Arg83 have changed in the lineage leading to mammals.

Discussion

The chicken gene BF2*1201 was the first classical class I gene identified outside of mammals (28, 29). Among the important findings from the structures of BF2*1201 and other analyses in this paper are three crucial observations: the identification of P2 as a promiscuous anchor residue, the role of Arg83 in allowing C-terminal overhangs out of the peptide-binding groove analogous to class II molecules, and the identification of Asp14 as an important conserved contact with β_2m, distinct from Arg14 in mammalian class I molecules.

Most classical class I molecules bind peptides with anchor residues at P2 (fitting into pocket B) and PΩ (fitting into pocket F), with only a few other configurations. One well-studied example is the mouse molecule H-2K^b, with obvious anchor residues at P5/6 and PΩ (9, 12). Subsequent studies showed that there is co-variation between P2 and P5/6, so that the residues at P2 are in fact constrained (13). For BF2*1201, sequences of both pooled and individual peptides showed mainly octamers with conserved amino acids at P5 and PΩ, primarily Val and Ile at P5 and Val at P8. A range of amino acids were found at P2, from the small hydrophobic Ala to the larger acidic Glu, with no obvious co-variation, so it was not clear whether P2 is an anchor residue.

475 The two structures of BF2*1201 go a long ways towards solving this mystery, with
476 Val2 and Pro2 clearly nestling under the α1 helix and bound into a canonical pocket B.
477 Thus, BF2*1201 has three anchor residues much like BF2*0401, but unlike BF2*0401,
478 P2 in BF2*1201 is promiscuous. Several other chicken class I molecules (BF2*0201,
479 1401 and 2101) show promiscuous binding of one kind or another, so perhaps this
480 should not have been a surprise. However, it remains to be determined how larger
481 residues like Glu and Gln at P2 interact with BF2*1201, perhaps by pocket B
482 accommodating only the Cβ and Cγ groups of the sidechains. Examination of a much
483 larger sample of bound peptides, for instance by immunopeptidomics or by assembly
484 using peptide libraries, might allow the possibility of co-variation to be conclusively
485 tested.
486
487 Perhaps the most striking observation from the BF2*1201 structures is the role of
488 Arg83 in allowing peptide overhang out of the C-terminal end of the groove. It was
489 noted very early on that highly conserved Tyr84 interacts with the peptide C-terminus
490 in mammals is replaced with Arg in the equivalent position of chickens and all other
491 non-mammalian vertebrates examined, and that this position is also an Arg in class II
492 molecules (33, 34). The significance of these observations had not been appreciated
493 until the two BF2*1201 structures showed that this Arg83 can interact with the
494 carboxyl group of the C-terminal amino acid in (at least) two conformations, one much
495 like Tyr84 with the C-terminal amino acid bound into pocket F and the other raised up
496 allowing the penultimate amino acid to bind into pocket F and the final amino acid to

extend out of the groove. This finding gives a molecular explanation to the suggestion that a nonamer peptide eluted from BF2*1201 might hang out of the groove, fits with unpublished structures for BF2*1401 (P. Chappell, M. Harrison, S. M. Lea and J. Kaufman, unpublished) and also fits with the motif found for a class I molecule from commercial chickens (N. Ternette, K. Watson and J. Kaufman, unpublished). Moreover, this finding can explain the unexpected peptide motif for BF2*0401 transfectants in cells mismatched for MHC haplotype, in which the usual Glu at PΩ was followed by a hydrophobic residue presumably specified by the alleles of TAP genes present (56). Different alleles of chicken class I molecules might have different propensities to allow these overhangs, and additional immunopeptidomics and structures may help understand this phenomenon further. In any case, it seems likely that such C-terminal overhangs will be found for all the non-mammalian vertebrates, allowing class I repertoires to be wider than is generally found in mammals.

It was shown long ago that a calreticulin peptide with a C-terminal overhang from the human class I molecule HLA-A2 led to a ten-fold loss of binding affinity, leading to the expectation that this was not a frequent occurrence (11). However, immunopeptidomics has identified many longer peptides (57-59). Some of these longer peptides may be due to extreme bulging in the middle (6, 60). However, other long peptides may involve C-terminal overhangs, as has been recently reported for two structures with a conformationally-shifted Tyr84, one with a *T. gondii* peptide bound to HLA-A*0201 and the other an insulin peptide bound to H-2K^{d} (49, 50).. These peptide overhangs

519 both involve significant shifts in the α-helices and Tyr84, but the conformational shift
520 of α-helices is not found in the corresponding structures of chicken class I, so the
521 binding mechanism is completely different from the simple conformational shift of
522 Arg83 in the chicken class I molecules. The shifted α-helices and Tyr84 to
523 accommodate C-terminally extended peptides may be more common as revealed by a
524 recent study in which peptidomics of 54 human class I alleles found C-terminal
525 extensions for at least eight alleles, including the common HLA-A03:01, HLA-A31:01,
526 and HLA-A68:01 (61). Moreover, the recent structures of H2-D^d and H2-D^b bound to
527 TAPBPR, the tapasin homolog, revealed the conformational shift of Tyr84 (62,63). In
528 contrast, the intrinsic Arg83 may suggest that the class I molecules from lower
529 vertebrates may have less need for either tapasin or TAPBPR during the peptide
530 loading.
531
532 It has long been believed that class I and class II molecules share a common ancestor,
533 although it has been controversial whether class I or class II molecules came first
534 (64-66). The fact that both class II molecules and chicken class I molecules allow
535 C-terminal peptide overhangs and that this correlates with the presence of a particular
536 Arg at the same position might be considered to support the idea that an open groove as
537 found in class II molecules came first. In fact, the conformations of the peptide
538 overhangs seem very similar (67). However, the conformation of Arg83 in chicken
539 class I and of the equivalent Arg in class II molecules is not the same; at the moment,
540 the evolutionary significance of this difference is not clear.

Finally, the appearance of Tyr84 in mammals instead of Arg83 in chickens and other non-mammalian vertebrates is not the only big change in the MHC I. For instance, the presence of the class III region between the regions encoding the class I and class II genes, the presence of functionally-monomorphic TAP and tapasin genes, and the presence of multigene families of well-expressed classical class I molecules all appeared in the lineage leading to the placental mammals (17, 36). Other changes might be expected, and one is the loss of Asp14 that is responsible for an intimate interaction between the α1 domain and β_2m. The significance of this change is not clear, but subtle differences in particular positions can have enormous effects. For instance, a Val to Ile change within the α3 domain abrogates CD8 binding in HLA*68:01 (68). Similarly, amino acid polymorphisms in mouse β_2m lead to big changes in T cell recognition (69). Thus, the conserved salt bridge of non-mammalian vertebrates that has been lost in mammals may have important consequences.

Author Contributions

W.J.L. and G.F.G. conceived and designed the experiments. J.X., W.X., Y.Z. W.P., M.Z. and L.N. performed the experiments. Y.C., J.Q., F.W. and W.J.L. analyzed the data. P.Q., C.P. and L.H. contributed reagents/materials/analysis tools. J.K, W.J.L., and G.F.G. contributed to the writing of the manuscript.

Acknowledgments

We thank the staff at the Shanghai Synchrotron Radiation Facility beamline 17U for assistance.

Disclosures

J.X., W.X., F.W..C.P., P.Q., and M.W. are employees of Zhongmu Institutes of China Animal Husbandry Industry Co. Ltd. These authors recognize the presence of a potential conflict of interest and affirm that the information represented in this paper is original and based on unbiased observations. This does not alter our adherence to journal policies on sharing data and materials. The other authors have no financial conflicts of interest.

References

1. Kelley, J., L. Walter, and J. Trowsdale. 2005. Comparative genomics of major histocompatibility complexes. *Immunogenetics* 56: 683-695.
2. Trowsdale, J., and J. C. Knight. 2013. Major histocompatibility complex genomics and human disease. *Annu. Rev. Genomics Hum. Genet.* 14: 301-323.
3. Rock, K. L., E. Reits, and J. Neefjes. 2016. Present Yourself! By MHC Class I and MHC Class II Molecules. *Trends Immunol.* 37: 724-737.
4. Stern, L. J., J. H. Brown, T. S. Jardetzky, J. C. Gorga, R. G. Urban, J. L. Strominger, and D. C. Wiley. 1994. Crystal structure of the human class II MHC protein HLA-DR1 complexed with an influenza virus peptide. *Nature* 368: 215-221.
5. Rammensee, H. G., K. Falk, and O. Rotzschke. 1993. Peptides naturally presented by MHC class I molecules. *Annu. Rev. Immunol.* 11: 213-244.

587 6. Bell, M. J., J. M. Burrows, R. Brennan, J. J. Miles, J. Tellam, J. McCluskey, J.
588 Rossjohn, R. Khanna, and S. R. Burrows. 2009. The peptide length specificity
589 of some HLA class I alleles is very broad and includes peptides of up to 25
590 amino acids in length. *Mol. Immunol.* 46: 1911-1917.
591 7. Saper, M. A., P. J. Bjorkman, and D. C. Wiley. 1991. Refined structure of the
592 human histocompatibility antigen HLA-A2 at 2.6 A resolution. *J. Mol. Biol.*
593 219: 277-319.
594 8. Guo, H. C., T. S. Jardetzky, T. P. Garrett, W. S. Lane, J. L. Strominger, and D. C.
595 Wiley. 1992. Different length peptides bind to HLA-Aw68 similarly at their
596 ends but bulge out in the middle. *Nature* 360: 364-366.
597 9. Matsumura, M., D. H. Fremont, P. A. Peterson, and I. A. Wilson. 1992.
598 Emerging principles for the recognition of peptide antigens by MHC class I
599 molecules. *Science* 257: 927-934.
600 10. Bouvier, M., and D. C. Wiley. 1994. Importance of peptide amino and carboxyl
601 termini to the stability of MHC class I molecules. *Science* 265: 398-402.
602 11. Collins, E. J., D. N. Garboczi, and D. C. Wiley. 1994. Three-dimensional
603 structure of a peptide extending from one end of a class I MHC binding site.
604 *Nature* 371: 626-629.
605 12. Fremont, D. H., M. Matsumura, E. A. Stura, P. A. Peterson, and I. A. Wilson.
606 1992. Crystal structures of two viral peptides in complex with murine MHC
607 class I H-2Kb. *Science* 257: 919-927.
608 13. Fremont, D. H., E. A. Stura, M. Matsumura, P. A. Peterson, and I. A. Wilson.

609 1995. Crystal structure of an H-2Kb-ovalbumin peptide complex reveals the
610 interplay of primary and secondary anchor positions in the major
611 histocompatibility complex binding groove. *Proc. Natl. Acad. Sci. USA* 92:
612 2479-2483.
613 14. Kaufman, J., S. Milne, T. W. Gobel, B. A. Walker, J. P. Jacob, C. Auffray, R.
614 Zoorob, and S. Beck. 1999. The chicken B locus is a minimal essential major
615 histocompatibility complex. *Nature* 401: 923-925.
616 15. Wallny, H. J., D. Avila, L. G. Hunt, T. J. Powell, P. Riegert, J. Salomonsen, K.
617 Skjodt, O. Vainio, F. Vilbois, M. V. Wiles, and J. Kaufman. 2006. Peptide motifs
618 of the single dominantly expressed class I molecule explain the striking
619 MHC-determined response to Rous sarcoma virus in chickens. *Proc. Natl. Acad.*
620 *Sci. USA* 103: 1434-1439.
621 16. Shaw, I., T. J. Powell, D. A. Marston, K. Baker, A. van Hateren, P. Riegert, M. V.
622 Wiles, S. Milne, S. Beck, and J. Kaufman. 2007. Different evolutionary
623 histories of the two classical class I genes BF1 and BF2 illustrate drift and
624 selection within the stable MHC haplotypes of chickens. *J. Immunol.* 178:
625 5744-5752.
626 17. Walker, B. A., L. G. Hunt, A. K. Sowa, K. Skjodt, T. W. Gobel, P. J. Lehner, and
627 J. Kaufman. 2011. The dominantly expressed class I molecule of the chicken
628 MHC is explained by coevolution with the polymorphic peptide transporter
629 (TAP) genes. *Proc. Natl. Acad. Sci. USA* 108: 8396-8401.
630 18. Tregaskes, C. A., M. Harrison, A. K. Sowa, A. van Hateren, L. G. Hunt, O.

631 Vainio, and J. Kaufman. 2016. Surface expression, peptide repertoire, and
632 thermostability of chicken class I molecules correlate with peptide transporter
633 specificity. *Proc. Natl. Acad. Sci. USA* 113: 692-697.
634 19. Kaufman, J. 2015. Co-evolution with chicken class I genes. *Immunol. Rev.* 267:
635 56-71.
636 20. Kaufman, J. 2015. What chickens would tell you about the evolution of antigen
637 processing and presentation. *Curr. Opin. Immunol.* 34: 35-42.
638 21. Kaufman, J. 2013. The avian MHC. In *Avian Immunology*, 2nd ed. K. A. Schat,
639 P. Kaiser, and B. Kaspers,eds. Academic Press. p. 149-167.
640 22. Miller, M. M., and R. L. Taylor, Jr. 2016. Brief review of the chicken Major
641 Histocompatibility Complex: the genes, their distribution on chromosome 16,
642 and their contributions to disease resistance. *Poult. Sci.* 95: 375-392.
643 23. Kaufman, J., H. Volk, and H. J. Wallny. 1995. A "minimal essential Mhc" and an
644 "unrecognized Mhc": two extremes in selection for polymorphism. *Immunol.*
645 *Rev.* 143: 63-88.
646 24. Kaufman, J. 2013. Antigen processing and presentation: evolution from a bird's
647 eye view. *Mol. Immunol.* 55: 159-161.
648 25. Koch, M., S. Camp, T. Collen, D. Avila, J. Salomonsen, H. J. Wallny, A. van
649 Hateren, L. Hunt, J. P. Jacob, F. Johnston, D. A. Marston, I. Shaw, P. R. Dunbar,
650 V. Cerundolo, E. Y. Jones, and J. Kaufman. 2007. Structures of an MHC class I
651 molecule from B21 chickens illustrate promiscuous peptide binding. *Immunity*
652 27: 885-899.

653 26. Zhang, J., Y. Chen, J. Qi, F. Gao, Y. Liu, J. Liu, X. Zhou, J. Kaufman, C. Xia,
654 and G. F. Gao. 2012. Narrow groove and restricted anchors of MHC class I
655 molecule BF2*0401 plus peptide transporter restriction can explain disease
656 susceptibility of B4 chickens. *J. Immunol.* 189: 4478-4487.
657 27. Chappell, P., K. Meziane el, M. Harrison, L. Magiera, C. Hermann, L. Mears, A.
658 G. Wrobel, C. Durant, L. L. Nielsen, S. Buus, N. Ternette, W. Mwangi, C. Butter,
659 V. Nair, T. Ahyee, R. Duggleby, A. Madrigal, P. Roversi, S. M. Lea, and J.
660 Kaufman. 2015. Expression levels of MHC class I molecules are inversely
661 correlated with promiscuity of peptide binding. *Elife* 4: e05345.
662 28. Guillemot, F., A. Billault, O. Pourquie, G. Behar, A. M. Chausse, R. Zoorob, G.
663 Kreibich, and C. Auffray. 1988. A molecular map of the chicken major
664 histocompatibility complex: the class II beta genes are closely linked to the
665 class I genes and the nucleolar organizer. *EMBO J.* 7: 2775-2785.
666 29. Kroemer, G., R. Zoorob, and C. Auffray. 1990. Structure and expression of a
667 chicken MHC class I gene. *Immunogenetics* 31: 405-409.
668 30. Butter, C., K. Staines, A. van Hateren, T. F. Davison, and J. Kaufman. 2013. The
669 peptide motif of the single dominantly expressed class I molecule of the chicken
670 MHC can explain the response to a molecular defined vaccine of infectious
671 bursal disease virus (IBDV). *Immunogenetics* 65: 609-618.
672 31. Reemers, S. S., D. A. van Haarlem, A. J. Sijts, L. Vervelde, and C. A. Jansen.
673 2012. Identification of novel avian influenza virus derived CD8+ T-cell
674 epitopes. *PLoS One* 7: e31953.

675 32. Hofmann, A., J. Plachy, L. Hunt, J. Kaufman, and K. Hala. 2003. v-src
676 oncogene-specific carboxy-terminal peptide is immunoprotective against Rous
677 sarcoma growth in chickens with MHC class I allele B-F12. *Vaccine* 21:
678 4694-4699.
679 33. Kaufman, J., R. Andersen, D. Avila, J. Engberg, J. Lambris, J. Salomonsen, K.
680 Welinder, and K. Skjodt. 1992. Different features of the MHC class I
681 heterodimer have evolved at different rates. Chicken B-F and beta
682 2-microglobulin sequences reveal invariant surface residues. *J. Immunol.* 148:
683 1532-1546.
684 34. Kaufman, J., J. Salomonsen, and M. Flajnik. 1994. Evolutionary conservation
685 of MHC class I and class II molecules--different yet the same. *Semin. Immunol.*
686 6: 411-424.
687 35. Kaufman, J. 1999. Co-evolving genes in MHC haplotypes: the "rule" for
688 nonmammalian vertebrates? *Immunogenetics* 50: 228-236.
689 36. Kaufman, J. 2011. The evolutionary origins of the adaptive immune system of
690 jawed vertebrates. In *The Immune Response to Infection*. S. H. E. Kaufmann, B.
691 T. Rouse, and D. L. Sachs,eds. American Society of Microbiology Press,
692 Washington, DC. p. 41-55.
693 37. Zhang, S., J. Liu, H. Cheng, S. Tan, J. Qi, J. Yan, and G. F. Gao. 2011. Structural
694 basis of cross-allele presentation by HLA-A*0301 and HLA-A*1101 revealed
695 by two HIV-derived peptide complexes. *Mol. Immunol.* 49: 395-401.
696 38. Otwinowski, Z., and W. Minor. 1997. Processing of X-ray diffraction data

collected in oscillation mode. *Methods Enzymol.* 276: 307-326.

39. Brunger, A. T., P. D. Adams, G. M. Clore, W. L. DeLano, P. Gros, R. W. Grosse-Kunstleve, J. S. Jiang, J. Kuszewski, M. Nilges, N. S. Pannu, R. J. Read, L. M. Rice, T. Simonson, and G. L. Warren. 1998. Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr. D Biol. Crystallogr.* 54: 905-921.

40. Emsley, P., and K. Cowtan. 2004. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* 60: 2126-2132.

41. Murshudov, G. N., A. A. Vagin, and E. J. Dodson. 1997. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D Biol. Crystallogr.* 53: 240-255.

42. Adams, P. D., R. W. Grosse-Kunstleve, L. W. Hung, T. R. Ioerger, A. J. McCoy, N. W. Moriarty, R. J. Read, J. C. Sacchettini, N. K. Sauter, and T. C. Terwilliger. 2002. PHENIX: building new software for automated crystallographic structure determination. *Acta Crystallogr. D Biol. Crystallogr.* 58: 1948-1954.

43. Laskowski, R. A., M. W. Macarthur, D. S. Moss, and J. M. Thornton. 1993. Procheck - a Program to Check the Stereochemical Quality of Protein Structures. *J. Appl. Crystallogr.* 26: 283-291.

44. Thompson, J. D., T. J. Gibson, F. Plewniak, F. Jeanmougin, and D. G. Higgins. 1997. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.* 25: 4876-4882.

719 45. Gouet, P., X. Robert, and E. Courcelle. 2003. ESPript/ENDscript: Extracting
720 and rendering sequence and 3D information from atomic structures of proteins.
721 *Nucleic Acids Res.* 31: 3320-3323.
722 46. Yao, S., J. Liu, J. Qi, R. Chen, N. Zhang, Y. Liu, J. Wang, Y. Wu, G. F. Gao, and
723 C. Xia. 2016. Structural Illumination of Equine MHC Class I Molecules
724 Highlights Unconventional Epitope Presentation Manner That Is Evolved in
725 Equine Leukocyte Antigen Alleles. *J. Immunol.* 196: 1943-1954.
726 47. Liu, J., Y. Sun, J. Qi, F. Chu, H. Wu, F. Gao, T. Li, J. Yan, and G. F. Gao. 2010.
727 The membrane protein of severe acute respiratory syndrome coronavirus acts as
728 a dominant immunogen revealed by a clustering region of novel functionally
729 and structurally defined cytotoxic T-lymphocyte epitopes. *J. Infect. Dis.* 202:
730 1171-1180.
731 48. Liu, W. J., J. Lan, K. Liu, Y. Deng, Y. Yao, S. Wu, H. Chen, L. Bao, H. Zhang, M.
732 Zhao, Q. Wang, L. Han, Y. Chai, J. Qi, J. Zhao, S. Meng, C. Qin, G. F. Gao, and
733 W. Tan. 2017. Protective T Cell Responses Featured by Concordant
734 Recognition of Middle East Respiratory Syndrome Coronavirus-Derived CD8+
735 T Cell Epitopes and Host MHC. *J. Immunol.* 198: 873-882.
736 49. McMurtrey, C., T. Trolle, T. Sansom, S. G. Remesh, T. Kaever, W. Bardet, K.
737 Jackson, R. McLeod, A. Sette, M. Nielsen, D. M. Zajonc, I. J. Blader, B. Peters,
738 and W. Hildebrand. 2016. Toxoplasma gondii peptide ligands open the gate of
739 the HLA class I binding groove. *Elife* 5: e12556.
740 50. Motozono, C., J. A. Pearson, E. De Leenheer, P. J. Rizkallah, K. Beck, A.

741 Trimby, A. K. Sewell, F. S. Wong, and D. K. Cole. 2015. Distortion of the Major
742 Histocompatibility Complex Class I Binding Groove to Accommodate an
743 Insulin-derived 10-Mer Peptide. *J. Biol. Chem.* 290: 18924-18933.
744 51. Liu, W. J., S. Tan, M. Zhao, C. Quan, Y. Bi, Y. Wu, S. Zhang, H. Zhang, H. Xiao,
745 J. Qi, J. Yan, W. Liu, H. Yu, Y. Shu, G. Wu, and G. F. Gao. 2016.
746 Cross-immunity Against Avian Influenza A(H7N9) Virus in the Healthy
747 Population Is Affected by Antigenicity-Dependent Substitutions. *J. Infect. Dis.*
748 214: 1937-1946.
749 52. Siddle, H. V., C. Sanderson, and K. Belov. 2007. Characterization of major
750 histocompatibility complex class I and class II genes from the Tasmanian devil
751 (Sarcophilus harrisii). *Immunogenetics* 59: 753-760.
752 53. Miska, K. B., G. A. Harrison, L. Hellman, and R. D. Miller. 2002. The major
753 histocompatibility complex in monotremes: an analysis of the evolution of Mhc
754 class I genes across all three mammalian subclasses. *Immunogenetics* 54:
755 381-393.
756 54. Stewart, R., Y. Ohta, R. R. Minter, T. Gibbons, T. L. Horton, P. Ritchie, J. D.
757 Horton, M. F. Flajnik, and M. D. Watson. 2005. Cloning and characterization of
758 Xenopus beta2-microglobulin. *Dev. Comp. Immunol.* 29: 723-732.
759 55. Wu, C., I. McConnell, and B. Blacklaws. 2008. Cloning and characterization of
760 ovine beta2-microglobulin cDNAs. *Vet. Immunol. Immunopathol.* 123:
761 360-365.
762 56. Sherman, M. A., R. M. Goto, R. E. Moore, H. D. Hunt, T. D. Lee, and M. M.

763 Miller. 2008. Mass spectral data for 64 eluted peptides and structural modeling
764 define peptide binding preferences for class I alleles in two chicken MHC-B
765 haplotypes associated with opposite responses to Marek's disease.
766 *Immunogenetics* 60: 527-541.
767 57. Giam, K., R. Ayala-Perez, P. T. Illing, R. B. Schittenhelm, N. P. Croft, A. W.
768 Purcell, and N. L. Dudek. 2015. A comprehensive analysis of peptides
769 presented by HLA-A1. *Tissue Antigens* 85: 492-496.
770 58. Schellens, I. M., H. D. Meiring, I. Hoof, S. N. Spijkers, M. C. Poelen, J. A. van
771 Gaans-van den Brink, A. I. Costa, H. Vennema, C. Kesmir, D. van Baarle, and C.
772 A. van Els. 2015. Measles Virus Epitope Presentation by HLA: Novel Insights
773 into Epitope Selection, Dominance, and Microvariation. *Front. Immunol.* 6:
774 546.
775 59. Trolle, T., C. P. McMurtrey, J. Sidney, W. Bardet, S. C. Osborn, T. Kaever, A.
776 Sette, W. H. Hildebrand, M. Nielsen, and B. Peters. 2016. The Length
777 Distribution of Class I-Restricted T Cell Epitopes Is Determined by Both
778 Peptide Supply and MHC Allele-Specific Binding Preference. *J. Immunol.* 196:
779 1480-1487.
780 60. Hassan, C., E. Chabrol, L. Jahn, M. G. Kester, A. H. de Ru, J. W. Drijfhout, J.
781 Rossjohn, J. H. Falkenburg, M. H. Heemskerk, S. Gras, and P. A. van Veelen.
782 2015. Naturally processed non-canonical HLA-A*02:01 presented peptides. *J.*
783 *Biol. Chem.* 290: 2593-2603.
784 61. Guillaume, P., S. Picaud, P. Baumgaertner, N. Montandon, J. Schmidt, D.E.

Speiser, G. Coukos, M. Bassani-Sternberg, P. Filippakopoulos, D. Gfeller. 2018. The C-terminal extension landscape of naturally presented HLA-I ligands. Proc Natl Acad Sci U S A. 115:5083-5088.

62. Jiang, J., K. Natarajan, L.F. Boyd, G.I. Morozov, M.G. Mage, D.H. Margulies. 2017. Crystal structure of a TAPBPR-MHC I complex reveals the mechanism of peptide editing in antigen presentation. Science. 358:1064-1068.

63. Thomas, C., R. Tampé. 2017. Structure of the TAPBPR-MHC I complex defines the mechanism of peptide loading and editing. Science. 358:1060-1064.

64. Kaufman, J. F., C. Auffray, A. J. Korman, D. A. Shackelford, and J. Strominger. 1984. The class II molecules of the human and murine major histocompatibility complex. *Cell* 36: 1-13.

65. Kaufman, J. 1988. Vertebrates and the evolution of the Major Histocompatibility Complex class I and class II molecules. *Verh. Dtsch. Zool. Ges.* 81: 131-144.

66. Flajnik, M. F., C. Canel, J. Kramer, and M. Kasahara. 1991. Which came first, MHC class I or class II? *Immunogenetics* 33: 295-300.

67. Kaufman, J. 2018. Unfinished business: evolution of the MHC and the adaptive immune system of jawed vertebrates. *Annu. Rev. Immunol.*: In press.

68. Salter, R. D., R. J. Benjamin, P. K. Wesley, S. E. Buxton, T. P. Garrett, C. Clayberger, A. M. Krensky, A. M. Norment, D. R. Littman, and P. Parham. 1990. A binding site for the T-cell co-receptor CD8 on the alpha 3 domain of HLA-A2.

Nature 345: 41-46.

69. Perarnau, B., C. A. Siegrist, A. Gillet, C. Vincent, S. Kimura, and F. A. Lemonnier. 1990. Beta 2-microglobulin restriction of antigen presentation. *Nature* 346: 751-754.

Figure legends

Figure 1. The binding capabilities of peptides to BF2*1201 revealed by in vitro refold. The refolding efficacy of B12 complex (BF2*1201, Chβ_2m and different epitopes) is evaluated by FPLC Superdex 200 16/60 Hi-Load gel-filtration (GE Healthcare). The peaks of the complexes with the expected molecular mass of 45 kDa were eluted at the estimated volume of 83 ml on a Superdex 200 column. The profile is marked with the approximate positions of the molecular mass standards of 75.0, 43.0 and 29.0 kDa. Peak ① represents the aggregated BF2*1201 heavy chain, peak ② represents the correctly refolded complex(42 kDa) and peak ③ represents excess chβ_2m. A. The peptides with a high capability to renature the BF2*1201 complex are shown. B. The peptides with no or minor capability to renature the BF2*1201 complex are shown.

Figure 2. The peptides in the narrow and hydrophobic groove of BF2*1201. (A and B) Conformations of peptides S1 (A) and A6 (B) in the groove of BF2*1201 are displayed through the 2*F*o-*F*c electron density maps contoured at 1.0 σ. The electron densities are shown as purple mesh viewed in profile with the α2 helix removed for clarity. (C and D) The P2, P5 and P8 anchors shown in sticks (green for S1 and cyan for A6) are located in the tangent plane of B, C and F pockets. The similar inserting depth of P2, P5 and P8 anchors in the pockets are shown with dashed purple lines linking the

Cα atoms of the residues in these three positions. (E and F) The vacuum electrostatic surface potential shows the A-F pockets of BF2*1201 labeled with letters in black circles. The peptides S1 and A6 are shown in green and cyan sticks.

Figure 3. BF2*1201 presents peptides in lower and flatter conformations compared with 8-mer peptides in MHC I from other species. S1 and A6 presented by BF2*1201 are represented by green sticks. All other 8-mer peptides, denoted as purple sticks, are presented by other mammalian MHC I alleles and chicken MHC I alleles. The lower position of the middle portion of peptides S1 and A6 presented by BF2*1201 denoted with blue dashed lines. The structures used are HLA-A*2402 (Protein Data Bank code 4F7T), HLA-B*3508 (3BWA), HLA-B*5101 (4MJI), H-2K^{b} (3P4M), and chicken BF2*0401 (4E0R).

Figure 4. Confirmation of the peptide motif for BF2*1201. (A) Binding of S1 and its Gly substitutions to BF2*1201 elucidated by *in vitro* refolding. (B) Capability of A6 and its Gly substitutions to renature BF2*1201. (C and D) Thermostabilities of BF2*1201 complexes. The thermostabilities of BF2*1201 with peptide S1 and A6 and their Gly substitutions (S1-P2G, S1-P5G, A6-P2G and A6-P9G) were tested by CD spectroscopy. The temperature was increased by 1°C/min. The curves for the unfolded fractions were determined by monitoring the CD value at 218 nm. Shown here are the data fitted to the denaturation curves using the Origin 8.0 program (OriginLab). The T_ms of different peptides are indicated by the gray line at 50% fraction unfolded.

Figure 5. The intrinsic open groove of BF2*1201. A. In the structure of BF2*1201/S1 (PDB:5YMW), The vacuum electrostatic surface potential shows the closed groove at the C-terminus of the peptide. B. Detailed interaction of P8-Val of peptide S1 with residues Arg81 and Thr141 of BF2*1201. C. The closed groove of HLA-A*0201/Mn2 (PDB:3I6G). D. The interaction of P9-Val of peptide Mn2 with Tyr83 and Thr141 of HLA-A*0201. E. The closed groove of H-2K^{d} (PDB: 5GSX). F. The interaction of P10-Leu with Tyr83 and Thr141 of H-2K^{d}. G. The opened groove at the C-terminus of peptide A6 presented by BF2*1201/S1 (PDB: 5YMV). H. The detailed interaction of hanging out residue P9-Val of peptide A6 with residues Arg81 and Thr141 of BF2*1201. I. The opened groove of HLA-A*0201/UFP(16–27) (PDB: 5DDH). J. No direct interaction of P12-Lys from peptide UFP(16–27) to Tyr83 can be observed. K. The opened groove of H-2K^{d}/G9GF (PDB: 4Z78). L. No direct interaction of the P10-Phe with residues Tyr81 and Thr141 of H-2K^{d}. M. The structural superposition of BF2*1201/S1 and BF2*1201/A6 shows no shift of the α-helices. N. The structural superposition of HLA-A*0201/Mn2 and HLA-A*0201/UFP(16–27) shows a conformational shift of the α1-helix of HLA-A*0201. O. The structural superposition of H-2K^{d}/142-2 and H-2K^{d}/G9GF highlights the conformational shift of the α1-helix of HLA-A*0201.

Figure 6. The intrinsic open groove enables BF2*1201 to bind to longer peptides with similar affinity. A. Binding of A6-C8 (AVKGVGTM) and the C-terminal

879 extended peptide A6 (AVKGVGTMV)to BF2*1201 elucidated by *in vitro* refolding.
880 B. Thermostabilities of BF2*1201 complexed to peptide A6 and A6-C8 examined by
881 CD spectroscopy. C. Binding of P9 (AIMEKNIVL) and the C-terminal extended
882 peptide P9-C10 (AIMEKNIVLK) to HLA-A*0201 elucidated by *in vitro* refolding. D.
883 Thermostabilities of HLA-A*0201 with peptide P9 and P9-C10 examined by CD
884 spectroscopy. E. Binding of P23 (RFSSFIRGKK) and the C-terminal extended peptide
885 P23-C11 (RFSSFIRGKKV) to HLA-A*1101 elucidated by *in vitro* refolding. F.
886 Thermostabilities of HLA-A*1101 with peptide P23 were tested by CD spectroscopy;
887 the C-terminal extended peptide (P23-C11) complexed to HLA-A*1101 was not
888 available due to the low binding.
889
890 **Figure 7.**
891
892 **Figure 8. Structure-based sequence alignment of BF2*1201 and other MHC I**
893 **molecules covering the residue at position 83.** Coils indicate α-helices, and black
894 arrows indicate β-strands. Residues highlighted in red are completely conserved, and
895 residues in blue boxes are highly (>80%) conserved, with consensus amino acids in red.
896 Residues at position 83 that are located at the C-terminus of the peptide binding groove
897 are marked with blue arrow. The residues at position 83 in lower vertebrate and chicken
898 are shown in yellow. The sequence alignment was generated with Clustal X and
899 ESPript.
900

Figure 9. The conserved Arg at the C-terminus of the peptide binding groove of MHC II molecules. A. Sequence alignment of MHC II molecules and other MHC I molecules for which structures are available covering the residue Arg83 of BF2*1201. The comparable residues at position 83 of BF2*1201 are marked with blue arrow. The conserved Arg residues corresponding to Arg83 of BF2*1201 that are located at the C-terminus of the peptide binding groove of MHC II are shown in yellow. B. Five representative MHC II structures (HLA-DR1, PDB:3L6F; HLA-DP2, PDB: 3LQZ; HLA-DQ6, PDB: 1UVQ; I-A^b, PDB: 1LNU; I-E^k, PDB: 1R5V) available in PDB were retrieved and superposed based on the α1 and β1 domains shown in helices and loops with different colors. The conserved Arg residues at the C-terminus of the peptide binding groove which are comparable to the Arg83 of BF2*1201 are shown in sticks. The MHC II-presented peptides with extended residues out of the grove were shown in ribbons with corresponding colors. C. The superposition of the five MHC II structures (shown in green) with BF2*1201/S1 (yellow), BF2*1201/A6 (cyan) and HLA-A*0201/Mn2 (purple, PDB: 3I6G). C. The superposition of BF2*1201/S1, BF2*1201/A6, HLA-A*0201/Mn2 (PDB: 3I6G) with other chicken MHC I molecules (BF2*0101, PDB: 4D0D; BF2*0401, PDB: 4E0R; BF2*1401, PDB: 4CW1; BF2*2101, PDB: 3BEW);.

E-I. Detailed C terminal anchoring mode of the five representative MHC II-presented peptides.